Appears in the Proceedings of the Fourteenth National Conference on Artificial Intelligence, July 1997, Providence, RI

# Knowledge Lean Word Sense Disambiguation*

**Ted Pedersen**
Department of Computer Science & Engineering
Southern Methodist University
Dallas, TX 75275–0122
pedersen@seas.smu.edu

A central problem in any natural language processing application is determining the meaning of ambiguous words. Word–sense disambiguation (WSD) is often cast as a problem in supervised learning. These approaches assume the availability of text to train a learning algorithm where ambiguous words have been manually annotated with sense distinctions. If such text is available, supervised approaches are effective and we present several extensions to an existing method. However, since sense–tagged text is expensive to create and only exists for a small number of ambiguous words, unsupervised alternatives are presented that do not require such an expensive knowledge source.

In both the supervised and unsupervised case, we approach WSD by assigning a sense to an ambiguous word based on a probabilistic model. The parametric form of a model describes the interactions between the features of a sentence with an ambiguous word. The estimates of the parameters of the model give the probability of observing all the possible combinations of feature values in a sentence.

Model selection is a supervised method of inducing a probabilistic model from sense–tagged text. We extend this approach in two ways. The search strategies and evaluation criteria used to select the parametric form of a probabilistic model for WSD are expanded and critically evaluated (Pedersen, Bruce, & Wiebe 1997). The Naive Mix, a new supervised learning algorithm that builds an averaged probabilistic model, is introduced and shown to be competitive with well–known machine learning algorithms (Pedersen & Bruce 1997).

In the absence of sense–tagged text, the sense of an ambiguous word is treated as a feature with a missing value. The observable features are those that can be automatically identified such as part–of–speech, morphology, and collocations. We perform WSD via two unsupervised learning algorithms, the EM algorithm and Gibbs Sampling. Both approaches assume that the parametric form is known; here we assume it to be the Naive Bayes model since it has been found to perform well in a wide range of problems including WSD.



The EM algorithm begins with guesses of the parameter estimates. The missing sense values are assigned expected values based on these estimates. The parameters are re–estimated based on these expected values. The sense values are reassigned based on the new parameter estimates. This process of re–estimating parameters and reassigning sense values iterates until the parameter estimates converge at a maxima. However, there is a danger that the estimates may find a local rather than global maxima, if the distribution of data is irregular.

Gibbs Sampling is an alternative to the EM algorithm that is based on Markov Chain Monte Carlo methods. It approximates the distribution of the estimates of each parameter though a sampling operation, whereas the EM algorithm finds point estimates via a maximization operation. Gibbs Sampling begins with guesses of the parameter estimates. The parameters are re–estimated by sampling from a conditional distribution where the previously sampled values of the other parameters are given as well as the values of the features. Sampling continues until an approximation for the distribution of the estimates for each parameter is obtained.

We cast WSD as a problem in unsupervised learning via the EM algorithm and Gibbs Sampling. This presents a significant advantage over supervised approaches in that it allows raw text to serve as the knowledge source rather than sense–tagged text.

## Acknowledgments

This research was supported by the Office of Naval Research under grant number N00014-95-1-0776.

## References

Pedersen, T., and Bruce, R. 1997. A new supervised learning algorithm for word sense disambiguation. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence*.

Pedersen, T.; Bruce, R.; and Wiebe, J. 1997. Sequential model selection for word sense disambiguation. In *Proceedings of the Fifth Conference on Applied Natural Language Processing*.